



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2018

Sprachtechnologie für Menschen mit Behinderungen

Ebling, Sarah

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-188575>

Conference or Workshop Item

Published Version



The following work is licensed under a Creative Commons: Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) License.

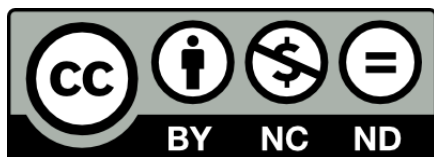
Originally published at:

Ebling, Sarah (2018). Sprachtechnologie für Menschen mit Behinderungen. In: Sonderpädagogik in der digitalisierten Lernwelt. Beiträge der nationalen Tagung Netzwerk Forschung Sonderpädagogik, Zurich, 4 September 2018. Edition SZH/CSPS, 29-46.

Empfohlene Zitierung

Ebling, S. (2018). Sprachtechnologie für Menschen mit Behinderungen. In S. Calabrese et al. (Hrsg./Éds.), *Sonderpädagogik in der digitalisierten Lernwelt. Beiträge der nationalen Tagung Netzwerk Forschung Sonderpädagogik. La pédagogie spécialisée dans l'environnement numérique d'apprentissage. Actes de la journée d'étude du Réseau de recherche en pédagogie spécialisée* (S. 29–46). Bern: Edition SZH/CSPS. Permalink: www.szh-csps.ch/b2018-01-02.

Dieser Artikel ist lizenziert unter einer [Creative Commons Namensnennung – Nicht kommerziell – Keine Bearbeitungen 4.0 International Lizenz](https://creativecommons.org/licenses/by-nc-nd/4.0/).



Sarah Ebling

Sprachtechnologie für Menschen mit Behinderungen

Abstract

An der Schnittstelle von Informatik und Sprachwissenschaft angesiedelt, beschäftigen sich die Computerlinguistik im Theorie- und die Sprachtechnologie im Anwendungsbereich mit der automatischen Verarbeitung natürlicher Sprache. Beispiele sind automatische Spracherkennung, Sprachsynthese, Sprachverstehen, Textanalyse, Sprachgenerierung oder Übersetzung. Die Sprachtechnologie wird im Kontext von Menschen mit Behinderungen in zwei Gebieten eingesetzt: einerseits in assistierenden Technologien und als Teil von E-Accessibility, andererseits in sonderpädagogischen Interventionen. Der vorliegende Beitrag diskutiert die Herausforderungen für die Sprachtechnologie in beiden Bereichen und skizziert davon ausgehend den Stand der Forschung.

Résumé

À l'intersection de l'informatique et de la linguistique, la linguistique informatique et la technologie linguistique se consacrent au traitement automatique du langage naturel - la première dans le domaine théorique, la seconde dans la pratique. La reconnaissance vocale, la synthèse vocale, la compréhension du langage, l'analyse de texte ou encore la traduction en sont des exemples. Dans le contexte du handicap, la technologie linguistique est employée dans deux domaines : d'une part, dans les technologies d'assistance et en tant qu'élément de l'e-accessibilité, d'autre part, dans des interventions de pédagogie spécialisée. La présente contribution examine les enjeux de la technologie linguistique dans ces deux domaines et donne à partir de là un aperçu de l'état actuel de la recherche.

Permalink: www.szh-csps.ch/b2018-01-02

Einleitung

Gus: *You're a really nice computer.*

Siri: *It's nice to be appreciated.*

Gus: *You are always asking if you can help me. Is there anything you want?*

Siri: *Thank you, but I have very few wants.*

Gus: *O. K.! Well, good night!*

Siri: *Ah, it's 5 : 06 p.m.*

Gus: *Oh sorry, I mean, goodbye.*

Siri: *See you later!*

Der eingangs aufgeführte Dialog erschien 2014 in einem Artikel der *New York Times*, in dem die Autorin beschreibt, wie ihr damals 13 Jahre alter Sohn mit einer Autismus-Spektrum-Störung Unterhaltungen mit Siri, der persönlichen Assistentin von Apple, führt.¹ Siri ist ein Beispiel für ein Dialogsystem; ein System, das mit menschlichen Benutzerinnen und Benutzern in eine Konversation tritt, um sie zu unterstützen, zu bilden oder zu unterhalten (Suendermann-Oeft, 2014).

Dialogsysteme sind Beispiele für Sprachtechnologien. Die Sprachtechnologie und ihre theoretische Schwesterdisziplin, die Computerlinguistik, sind an der Schnittstelle von Informatik und Sprachwissenschaft angesiedelt und beschäftigen sich mit der automatischen Verarbeitung natürlicher Sprachen. Dies schliesst sowohl Lautsprachen als auch Gebärdensprachen ein. Beispiele für Sprachtechnologien beziehungsweise Computerlinguistik-Anwendungen sind

- die automatische Spracherkennung, bei der gesprochene oder gebärdete Sprache maschinell in geschriebene Sprache umgewandelt wird,
- die automatische Sprachsynthese, bei der aus geschriebener Sprache maschinell gesprochene oder gebärdete Sprache erzeugt wird,
- das automatische Sprachverstehen, bei dem natürlichsprachlicher Text maschinell in eine formale Bedeutungsrepräsentation überführt wird (vgl. Bsp. 1 als Möglichkeit einer solchen Repräsentation für den Satz «I want to go from Chicago to New York on August 17 in the early afternoon on United.»),
- die automatische Textanalyse, bei der ein Text auf verschiedenen linguistischen Ebenen (etwa der Ebene der Morphologie, Morphosyntax oder Syntax) maschinell analysiert wird,

¹ Newman (2014).

- die automatische Sprachgenerierung, bei der eine formale Bedeutungsrepräsentation maschinell in natürlichsprachlichen Text konvertiert wird, und
- die maschinelle Übersetzung, bei der automatisch von einer schriftlichen Repräsentation einer Quellsprache (Lautsprache oder Gebärdensprache) in eine schriftliche Repräsentation einer Zielsprache (Lautsprache oder Gebärdensprache) übersetzt wird (vgl. Bsp. 2 für eine Übersetzung von Deutsch in eine schriftliche Form – Glossen – der Deutschen Gebärdensprache).

Beispiel 1

Formale Bedeutungsrepräsentation des Satzes «I want to go from Chicago to New York on August 17 in the early afternoon on United.» (Beals et al., 2015)

Destination: New York

Departure city: Chicago

Departure date: August 17

Departure time: early afternoon

Airline: United

Beispiel 2

Übersetzung von Deutsch in Glossen der Deutschen Gebärdensprache (Stein, Schmidt & Ney, 2012).

Die Temperaturen sinken in der Nacht auf 11 Grad an der Nordsee und 4 Grad an den Alpen. → *TEMPERATUR NACHT SINKEN 11 NORDEN SEE 4 GRAD ALPEN*²

Sprachtechnologie enthält heutzutage einen hohen Anteil an künstlicher Intelligenz. Verbindungen von Sprachtechnologie und künstlicher Intelligenz ergeben sich etwa über das maschinelle Lernen. Das Grundprinzip maschinellen Lernens besteht darin, dass ein System auf der Basis existierender Beispiele lernt und unter Zuhilfenahme von Statistik neue Aufgaben löst. Für die bereits erwähnte maschinelle Übersetzung bedeutet dies zum Beispiel, dass von menschlichen Expertinnen und Experten angefertigte Übersetzungen

² Die gebärdensprachliche Übersetzung ist hier in Glossen notiert. Glossen sind semantische Bezeichnungen für Gebärden, die üblicherweise die Grundform eines Wortes der umgebenden Lautsprache annehmen und grossgeschrieben werden, um anzuzeigen, dass es sich um Referenzen auf gebärdensprachliche Konzepte handelt.

gen in Teilen oder als Ganzes als Beispiele dienen, um neue Übersetzungen zu generieren. Die Gesamtheit solcher verwendeten Beispiele wird als «Trainingsmaterial» bezeichnet.

In der Spracherkennung als eine der oben eingeführten Sprachtechnologien ist eine wichtige Unterscheidung die Zusammensetzung des Trainingsmaterials: Bei der *sprecherabhängigen* Spracherkennung wird eine benutzende Person aufgefordert, vorgängig einige Sätze in die Software zu sprechen, die dann dem System als zusätzliches Trainingsmaterial dienen. Diese Art der Spracherkennung fördert in der Regel die besten Ergebnisse zutage. Bei der *sprecherunabhängigen* Spracherkennung findet kein zusätzliches benutzerspezifisches Training statt. Entsprechend sind die Ergebnisse von tieferer Qualität. Eine Zwischenform stellt die *sprecheradaptive* Spracherkennung dar, bei der kein initiales Training, jedoch ein Anpassungsvorgang während der Benutzung stattfindet.

Die eingangs beschriebene Verwendung von Siri als Dialogsystem durch einen Jungen mit einer Autismus-Spektrum-Störung ist ein Beispiel für den Einsatz von Sprachtechnologie im Kontext von Menschen mit Behinderungen und besonderem Bildungsbedarf. An dieser Schnittstelle ist der vorliegende Beitrag angesiedelt: Er hat zum Ziel, die Herausforderungen, vor die einzelne Sprachtechnologien im genannten Kontext gestellt sind, zu diskutieren und davon ausgehend den Stand der Forschung zu skizzieren.

Liesen und Rummeler (2016, S. 7) definieren drei Dimensionen des Einbezugs von *Information and Communication Technology* (ICT) bzw. digitalen Medien in der Sonder- und Medienpädagogik: Barrierefreiheit (*Accessibility*), Assistieren mit ICT und Fördern mit ICT. Für die Sprachtechnologie lassen sich zwei Einsatzgebiete im Kontext von Menschen mit Behinderungen ansetzen, auf die im Folgenden näher eingegangen wird:

1. Sprachtechnologie in assistierenden Technologien und als Teil von *E-Accessibility*
2. Sprachtechnologie in sonderpädagogischen Interventionen

Sprachtechnologie in assistierenden Technologien und als Teil von E-Accessibility

Sprachtechnologie findet sich einerseits in technischen Hilfsmitteln, die Menschen mit Behinderungen und älteren Menschen den Zugang zu ICT und damit zu elektronisch übertragener Information und Kommunikation ermöglichen (assistierende Technologien). Andererseits kann sie auch Teil der Her-

stellung von Zugänglichkeit zu ICT auf Anbieterseite sein (*E-Accessibility*). Hierzu gehören die Herstellung barrierefreier digitaler Dokumente (z. B. Webseiten und PDFs) und Software sowie die *Access Services* der audiovisuellen Medien (z. B. Untertitelung, Audiodeskription).³ Im Folgenden werden beispielhaft Anwendungen von Sprachtechnologien vorgestellt, die Teil von assistierenden Technologien oder *E-Accessibility* sind.

Automatische Spracherkennung

Zum Auftrag des Schweizer Fernsehens gehört es, einen bestimmten Prozentsatz der ausgestrahlten Sendungen mit Untertiteln für Zuschauerinnen und Zuschauer mit Hörbehinderungen zu versehen. Verantwortlich für die Herstellung von Untertiteln ist SWISS TXT, ein Tochterunternehmen der SRG SSR. Bei der Produktion von Live-Untertiteln setzt SWISS TXT *Respeaking* ein, eine Technik, bei der menschliche Untertitlerinnen oder Untertitler die Redebeiträge eines Live-Segments unter Einbezug von Interpunktion und zusätzlicher Information in eine Spracherkennungssoftware nachsprechen (Romero-Fresco, 2011).

Automatische Spracherkennung findet sich auch als maschinelle Erkennung dysarthrischen⁴ Sprechens in elektronischen Hilfsmitteln der Unterstützten Kommunikation (UK). Bei leichter bis mittelschwerer Dysarthrie können hier mit gebrauchsfertigen Spracherkennungssystemen zufriedenstellende Ergebnisse erzielt werden. Bei schwerer Dysarthrie ist in der Regel eine angepasste Spracherkennungslösung notwendig. Hier ergeben sich vor allem zweierlei Herausforderungen (Hawley et al., 2013; Rudzicz, 2016): Zum einen steht diesen Systemen vergleichsweise wenig Trainingsmaterial zur Verfügung, weil es für Menschen mit Dysarthrie eine grosse Anstrengung bedeutet, Sätze für das Training eines Systems zu sprechen. Zum anderen weisen wiederholte Sätze, die für das Training des Systems notwendig sind, bei Menschen mit Dysarthrie üblicherweise eine grössere Varianz auf als bei nicht dysarthrischen Personen. Diesem Umstand wird in der Forschung Rechnung getragen, indem Datenbanken mit Beispielen dysarthrischen

³ Vollständige digitale Barrierefreiheit ist zum heutigen Zeitpunkt kaum möglich. Entsprechend müsste richtigerweise von *barrierearmen* digitalen Dokumenten gesprochen werden. Da sich der Begriff *Barrierefreiheit* als Übersetzung des englischen Begriffs *Accessibility* im Deutschen aber weitgehend eingebürgert hat, wird er aus Konsistenzgründen auch in diesem Beitrag verwendet.

⁴ Die Dysarthrie ist ein Sammelbegriff für verschiedene sprechmotorische Störungen, die aufgrund einer Schädigung des zentralen Nervensystems auftreten.

Sprechens aufgebaut werden, die dann über Forschungsgruppen hinweg verfügbar sind. Ein Beispiel hierfür ist etwa die TORGO-Datenbank für den englischsprachigen Raum (Rudzicz, Namasivayam & Wolff, 2012).

Auch im Kontext von *Active Assisted Living* (AAL, vormals *Ambient Assisted Living*), der Gesamtheit von Technologien, die es Menschen mit Behinderungen und älteren Menschen erlauben, länger in ihrem Zuhause zu leben, kommt Spracherkennung zur Anwendung. Sie erfüllt dort die Funktion einer Kommandoschnittstelle, ermöglicht Benutzerinnen und Benutzern also das Absetzen von festgelegten Befehlen, etwa zum Tätigen eines Notrufs, Öffnen und Schliessen von Fensterläden oder Ein- und Ausschalten von Lichtschaltern. Für die Spracherkennungskomponente stellen sich im AAL-Kontext einige zentrale Herausforderungen (Vacher et al., 2013): Zunächst handelt es sich um Spracherkennung auf Distanz, was das Vorhandensein von Umgebungsgeräuschen impliziert. Hinzu kommt, dass die Stimmen von Benutzerinnen und Benutzern im AAL-Kontext (Menschen mit Behinderungen oder älteren Menschen) spezifische Merkmale aufweisen können, sodass eine angepasste Spracherkennungslösung erforderlich wird. Das System muss zudem ständig einsatzbereit sein, ohne jedoch die Privatsphäre des Benutzers oder der Benutzerin zu stören. In der Praxis wird diesem Spannungsfeld meist begegnet, indem die Spracherkennungskomponente erst durch ein Schlüsselwort aktiviert wird. Zudem sind die Spracherkennungskomponenten oft bewusst mit einem eingeschränkten Vokabular ausgestattet, das nur die für den AAL-Kontext relevanten Aufgaben abdeckt.

Spracherkennung kann auch in der visuell-räumlichen Modalität von Gebärdensprachen stattfinden. Hierbei wird Gebärdensprache zumeist auf Basis eines Videos erkannt und in eine schriftliche Repräsentation überführt, die sowohl die Form als auch die Bedeutung der erkannten Gebärden beschreibt. Ein Beispiel einer Formverschriftlichung stellt das Hamburger Notationssystem für Gebärdensprachen (HamNoSys) dar (Prillwitz et al., 1989). Abbildung 1 zeigt die HamNoSys-Notation zur Gebärde für das Konzept *Volk* in der Deutschschweizerischen Gebärdensprache (DSGS). Gebärdenbedeutungen werden meist mittels Glossen notiert (z. B. VOLK, vgl. auch Bsp. 2).

Die automatische Gebärdenspracherkennung verläuft in zwei Schritten: Zuerst werden für das Gebärden relevante Positionsmerkmale extrahiert, danach wird deren zeitlicher Verlauf modelliert. Zu den Herausforderungen des maschinellen Sehens (*Computer Vision*) gehören etwa Verdeckungen, schwierige Lichtverhältnisse und ungewohnte Perspektiven. Hinzu kommt die gebärdensprachspezifische Herausforderung der Erkennung mehrerer

Produktionsmodalitäten, das heisst von Händen und Armen (manuellen Komponenten) und Gesicht, Kopf und Schultern (nicht-manuellen Komponenten). Zuverlässige Ergebnisse liefert die automatische Gebärdenspracherkennung bis dato nur auf der Ebene der Einzelgebärden und für eingeschränkte Vokabulare (Ebling, Camgöz & Bowden, im Erscheinen).

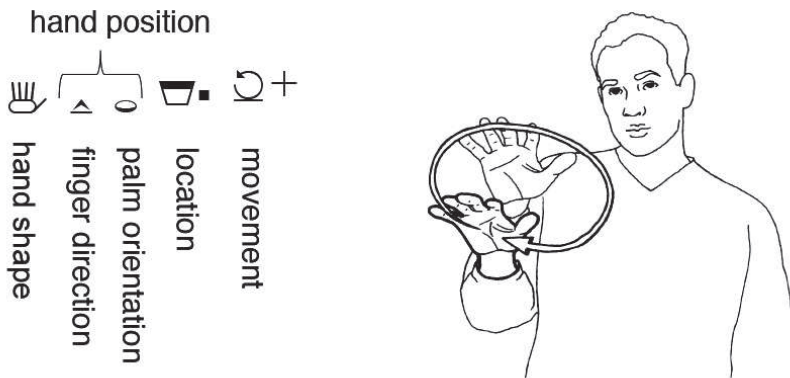


Abbildung 1: HamNoSys-Notation und Illustration der Gebärde VOLK in Deutschschweizerischer Gebärdensprache (Illustration: Boyes Braem, 2001)

Automatische Sprachsynthese

Eine bekannte Sprachsynthese-Anwendung innerhalb der assistierenden Technologien stellen Screenreader dar, die Blinden und Menschen mit einer Sehbehinderung die Interaktion mit dem Computer ermöglichen, indem sie ihnen textuell repräsentierte Inhalte vorlesen. Blinde und Menschen mit einer Sehbehinderung wählen für ihre Screenreader meist eine hohe Sprechgeschwindigkeit, um in der Lage zu sein, schnell einen akustischen Überblick über ein elektronisches Dokument zu erhalten (etwa eine Webseite oder ein PDF). Die Möglichkeit, die Sprechgeschwindigkeit individuell zu variieren, ist auch der Grund, weshalb in Screenreadern kein Syntheseverfahren zur Anwendung kommt, das dem Stand der Forschung entspricht. Stattdessen basieren Screenreader auf der *Formantensynthese*, welche die Eigenschaft hat, auch bei hoher Sprechgeschwindigkeit gute Ergebnisse zu erzeugen.

Die automatische Sprachsynthese mit ihren aktuellsten Verfahren findet Eingang in den Prozess der Herstellung von Audiodeskriptionen, das

heisst von akustischen Darstellungen visueller Informationen beispielsweise im Fernsehen (wo die Bezeichnung *Hörfilm* geläufig ist). Durch den Einbezug von Sprachsynthese entfällt die Vertonung des Audiodeskriptions-Skripts in einem Studio mit einem professionellen Sprecher oder einer professionellen Sprecherin, bei der auch eine blinde Person für das *proof listening* anwesend sein muss. Dieses *proof listening* kann dank der synthetisierten Vertonung der Audiodeskription orts- und zeitunabhängig stattfinden. Das Schweizer Fernsehen sendet regelmässig synthetisierte Audiodeskriptionen.

Auch Gebärdensprache lässt sich künstlich erzeugen. Die automatische Gebärdensprachsynthese, bei der aus einer schriftlichen Repräsentation von Gebärdensprache, etwa der in Abbildung 1 dargestellten HamNoSys-Notation, ein Gebärdensprach-Avatar erzeugt wird, ist abzugrenzen von Gebärdensprach-Animationen, die durch aufwändiges Modellieren und Animieren von virtuellen *Characters* in einer Animations-Software erstellt werden (*Animation von Hand*). Solcherart produzierte Animationen sind zwar von hoher Qualität, eignen sich jedoch nicht für die Verwendung als Teil eines ICT-Systems, wie es in Abbildung 2 dargestellt ist: Hier dient die Ausgabe eines Übersetzungssystems als Eingabe für einen Gebärdensprach-Avatar. Dies bedeutet, dass der Gebärdensprach-Avatar mit einer beliebigen, nicht vorhersehbaren Eingabe umgehen können muss. Diese Eigenschaft bietet nur die vollautomatische Gebärdensprachsynthese, bei der sich alle Gebärden erzeugen lassen, die durch ein zugrundeliegendes Notationssystem (etwa HamNoSys) beschrieben werden können. Dieses vollautomatische Verfahren liefert gegenüber der Animation von Hand eine tiefere Qualität (Ebling, 2013; Kipp et al., 2011). Vollständig synthetisierte Gebärdensprach-Avatare können in Verbindung mit Gebärdensprachübersetzung beispielsweise eingesetzt werden, um öffentliche Informationen (etwa Zugansagen) zu übermitteln; ihr Einsatz in unrestringierten Domänen, beispielsweise auf beliebigen Webseiten, ist derzeit noch nicht realistisch. Forschungsbedarf existiert vor allem im Bereich der Darstellung nicht-manueller Merkmale (Kipp, Heloir & Nguyen, 2011).

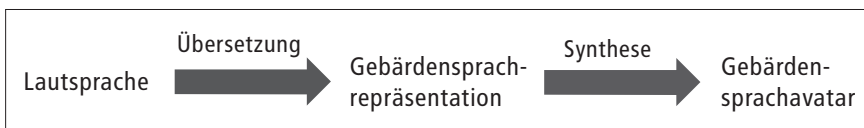


Abbildung 2: Transfer von Lautsprache in synthetisierte Gebärdensprache:
Übersetzung und Synthese (Ebling, 2016)

Maschinelle Übersetzung

Die zunehmende Verfügbarkeit von Informationen in Leichter Sprache ermöglicht die automatische Textvereinfachung auf der Basis maschinellen Lernens. In jüngster Zeit wird diese als Aufgabe der maschinellen Übersetzung aufgefasst. Dies setzt das Vorhandensein von *parallelen Korpora* voraus, von Sammlungen, in denen Sätze in einer Sprache wie Deutsch oder Englisch Sätzen in vereinfachtem Deutsch oder vereinfachtem Englisch zugeordnet sind. Fürs Englische wird hierfür typischerweise die Kombination aus *English Wikipedia* und *Simple English Wikipedia* verwendet (Coster & Kauchak, 2011).⁵ In Ermangelung einer vereinfachten Form der deutschen Wikipedia muss für diese Sprache auf andere Datenquellen ausgewichen werden. Mittlerweile existieren zahlreiche Webseiten, die Informationen sowohl in Deutsch als auch in leichtem Deutsch anbieten. Klaper, Ebling und Volk (2013) erstellten daraus ein paralleles Korpus als Datenbasis für ein maschinelles Übersetzungssystem.

Der Begriff *Gebärdensprachübersetzung* bezeichnet entweder die Übersetzung von schriftlicher Lautsprache in schriftliche Gebärdensprache, wie sie in Beispiel 2 dargestellt ist, die Übersetzung in die entgegengesetzte Richtung oder die Übersetzung von einer Gebärdensprache in eine andere (in schriftlicher Form). Da Gebärdensprache in schriftlicher Form für gehörlose Gebärdensprachbenutzende keinen praktischen Nutzen aufweist, muss Gebärdensprachübersetzung mit Gebärdenspracherkennung oder Gebärdensprachsynthese kombiniert werden, um Teil einer ICT-Lösung zu sein. Abbildung 2 zeigt die Kombination von Gebärdensprachübersetzung mit Gebärdensprachsynthese. Abbildung 3 bringt Gebärdenspracherkennung und -übersetzung zusammen, und Abbildung 4 schliesslich kombiniert alle drei Gebärdensprachtechnologien (-erkennung, -übersetzung und -synthese).

⁵ https://en.wikipedia.org/wiki/Main_Page [Zugriff am 05.10.2018], https://simple.wikipedia.org/wiki/Main_Page [Zugriff am 05.10.2018].

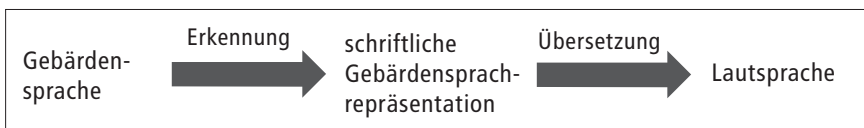


Abbildung 3: Transfer von Gebärdensprache in Lautsprache (Ebling, 2016)

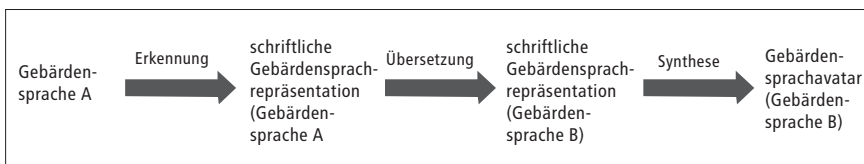


Abbildung 4: Transfer von einer Gebärdensprache in eine synthetisierte Form einer anderen Gebärdensprache (Ebling, 2016)

Sprachtechnologie in sonderpädagogischen Interventionen

Sprachtechnologie ist imstande, einen Beitrag zu Interventionen in der Sonderpädagogik zu leisten, womit der gesamte Prozess von der Diagnostik über die eigentliche Intervention bis hin zur Evaluation gemeint ist. Der Beitrag der Sprachtechnologie konzentriert sich hierbei vor allem auf die Diagnostik und die eigentliche Intervention, d. h. die Förderung oder Therapie. Im Folgenden werden Anwendungen einiger Sprachtechnologien in sonderpädagogischen Interventionen skizziert.

Automatische Spracherkennung

Eine Anwendung der Spracherkennung, die der eigentlichen Diagnostik vorgelagert ist, bildet die automatische Transkription von Spontansprachbeispielen. Sprachbeispiele werden hier also automatisch verschriftlicht, um sie überhaupt erst systematisch analysieren zu können.

Automatische Spracherkennung kommt auch in der Dysarthrie- (Frieg et al., 2017; Ganzeboom et al., 2016; Krause, Smeddinck & Meyer, 2013; Martens et al., 2015) und Aphasietherapie (Beals et al., 2015) zum Einsatz, wobei es sich genau genommen meist lediglich um Signalerkennung und keine eigentliche Spracherkennung handelt: Erkannt werden beispielsweise Merkmale wie Lautstärke oder Tonhöhe, nicht Wörter und Sätze. Bisweilen sind die Erkennungskomponenten Teil eines *Serious Game*, bei dem der Benutzer

oder die Benutzerin eine bestimmte Aufgabe löst. Das Spiel *Treasure Hunters* (Ganzeboom et al., 2016) für das Niederländische beispielsweise ist als Ergänzung zu einer Sprachtherapie konzipiert und stellt zwei Spielende, eine Person mit Dysarthrie und eine Person aus deren Bekanntenkreis, vor die Aufgabe, einen Schatz und den dazugehörigen Schlüssel zu finden, indem sie sich über eine Audioverbindung miteinander unterhalten. Das Programm erkennt Lautstärke und Tonhöhe der beiden Sprecherinnen oder Sprecher und liefert Rückmeldungen dazu.

Automatische Textanalyse

Sind Spontansprachbeispiele einmal transkribiert, das heisst (mit oder ohne automatische Spracherkennung) verschriftlicht, kann automatische Textanalyse auf unterschiedlichen linguistischen Ebenen angewendet werden, etwa der morphologischen und der syntaktischen, um im Rahmen der Diagnostik sprachliche Hinweise etwa auf eine spezifische Sprachentwicklungsstörung (Solorio, 2013) oder eine leichte kognitive Beeinträchtigung (Lehr et al., 2012) zu erhalten. Textanalysevorgänge dieser Art basieren auf der automatischen Klassifikation als Unterverfahren des maschinellen Lernens. Hierbei werden Klassifikationsmerkmale wie beispielsweise strukturelle Komplexität, semantische und syntaktische Fehler, Sprechflüssigkeit, Kohärenz oder narrative Performanz berücksichtigt (Solorio, 2013). Auch bei der Bestimmung des Sprachentwicklungsstandes im syntaktisch-morphologischen Bereich wird automatische Textanalyse eingesetzt (Sagae, Lavie & MacWhinney, 2005). Mit Hilfe automatischer Textanalyse können zudem Texte für die Förderung bzw. Therapie oder für den Unterricht gefunden werden, die ein bestimmtes linguistisches Komplexitätsniveau aufweisen. Eine Analyse dieser Art steht in der Tradition der Lesbarkeitsmasse, berücksichtigt jedoch linguistisch komplexere und damit aussagekräftigere Merkmale als die für Lesbarkeitsmasse typischen Indikatoren Wort- und Satzlänge (Heimann Mühlenbock, 2013).

Automatische Textanalyse ist auch Bestandteil eines *Proof of concept*, der im Rahmen des Projekts «Computergestütztes Übungsprogramm für psychomotorische Diagnostik» an der Interkantonalen Hochschule für Heilpädagogik Zürich (HfH) erstellt wurde.⁶ Entwickelt wurde der Prototyp eines Online-Tools, das es Studierenden der Psychomotoriktherapie erlaubt, ihre diagnostischen Texte maschinell mit jenen von Expertinnen und Experten abzugleichen. Der Abgleich findet auf der Bedeutungsebene statt (semantische Ähnlichkeit).

Automatisches Sprachverstehen

Die automatische Textanalyse, wie sie im vorangegangenen Kapitel vorgestellt wurde, ist Teil des automatischen Sprachverstehens, bei dem natürlichsprachlicher Text in eine formale Bedeutungsrepräsentation überführt wird (vgl. Bsp. 1). Ein Beispiel für ein Programm im Therapiebereich, das automatisches Sprachverstehen beinhaltet, ist *Grammar Trainer* (Beals et al., 2015). Es richtet sich an Kinder und Jugendliche mit einer Autismus-Spektrum-Störung. Das Programm präsentiert der Benutzerin oder dem Benutzer eine Situation, meist unter Zuhilfenahme von Illustrationen, und eine zugehörige Frage in schriftlicher Form (aufseiten der Benutzerin oder des Benutzers wird also eine grundlegende Lesekompetenz vorausgesetzt). Ein Beispiel einer solchen Vignette ist die visuelle und textuelle Darstellung eines Jungen, der gegenüber einem Mädchen den Satz äußert: «Come over to my house and play with me.» Die Benutzerin oder der Benutzer ist angehalten, eine Freitextantwort auf die Frage «What will the boy want?» zu geben, und er hält daraufhin vom System eine Rückmeldung auf den Antwortsatz, der für das vorliegende Beispiel korrekterweise «He will want the girl to come over to his house and play with him» lautet. Eine Rückmeldung kann sich etwa auf das Fehlen eines Artikels beziehen, auf eine falsche Wortformenendung oder auf das Fehlen eines Inhaltswortes. Um derartige Rückmeldungen zu erzeugen, führt das System intern eine automatische morphologische, morphosyntaktische, syntaktische und semantische Analyse durch.

Dialogsysteme

Automatisches Sprachverstehen ist Bestandteil von Dialogsystemen, wie sie zu Beginn dieses Beitrags eingeführt wurden, und bildet zusammen mit Dialogmanagement und automatischer Sprachgenerierung die drei Kernschritte eines Dialogsystems *geschriebener Sprache*. Je nach Modalität wird dieser Kern umrahmt von zusätzlichen Sprachtechnologien: Bei den gängigsten Dialogsystemen, diejenigen *gesprochener Sprache*, wird der Prozess eingeleitet durch automatische Spracherkennung und abgeschlossen durch automatische Sprachsynthese. Derartige Systeme kommen beispielsweise in Telefon-Hotlines oder als Teil der eingangs eingeführten persönlichen Assistenten zum Einsatz (z. B. Siri, Alexa oder Cortana). Multimodale Dialogsysteme zeichnen

⁶ https://www.hfh.ch/de/forschung/projekte/computergestuetztes_uebungsprogramm_fuer_psychomotorische_diagnostik/ [Zugriff am 05.10.2018].

sich dadurch aus, dass sie neben Spracherkennung und Sprachsynthese auch eine automatische Erkennung beispielsweise von Gesten, Gesichts- oder Körperausdruck sowie eine automatische Ausgabe von visueller Information beinhalten. *Embodied Conversational Agents* (ECA), Dialogsysteme mit grafisch modellierten und animierten *Characters*, sind Beispiele für multimodale Dialogsysteme. ECA, bei denen nur der Kopf dargestellt ist, werden als *Talking Heads* bezeichnet. ECA finden sich in der Sonderpädagogik beispielsweise in der Therapie von Menschen mit einer Autismus-Spektrum-Störung (Bosseler & Massaro, 2003; Milne et al., 2011; Smith et al., 2014; Tartaro & Cassell, 2008) oder mit Aphasie (Pompili et al., 2015; Teodoro et al., 2013). Zu bedenken ist, dass ECA gegenüber Dialogsystemen geschriebener und gesprochener Sprache eine höhere kognitive Beanspruchung mit sich bringen (Lopez-Cozar et al., 2011; Veletsianos & Russell, 2013).

Zusammenfassung und Ausblick

Der vorliegende Artikel hat den Beitrag der Computerlinguistik zu den Bereichen assistierende Technologien respektive *E-Accessibility* und sonderpädagogische Interventionen aufgezeigt. Thematisiert wurden computerlinguistische Anwendungen und Sprachtechnologien wie automatische Spracherkennung, Sprachsynthese, Sprachverstehen, Textanalyse, Sprachgenerierung und Übersetzung. Einzelne der vorgestellten Anwendungen, etwa die automatische Erkennung dysarthrischen Sprechens, können sowohl die Funktion assistierender Technologien als auch jene von Förder- oder Therapietools übernehmen.

Hinsichtlich des Beitrags der Computerlinguistik und der Sprachtechnologie zu den genannten Gebieten ist die vorgestellte Übersicht keineswegs vollständig. Weitere Anwendungsbeispiele sind etwa die

- automatische Sprachsynthese in *Talkern* (Sprachcomputern) im UK-Kontext
- Wortvorhersage und -vervollständigung im UK-Kontext,
- optische Zeichenerkennung (*optical character recognition, OCR*) als Vorverarbeitungsschritt, um Text, der als Bild abgespeichert ist, einem Screenreader zugänglich zu machen, und die
- automatische Erzeugung von Bildbeschreibungen als Alternativtexte für Screenreader in digitalen Dokumenten.

Der Beitrag hat verdeutlicht, dass einige Ansätze im Kontext von Menschen mit Behinderungen stark interdisziplinär angelegt sind. So setzt beispiels-

weise die automatische Gebärdenspracherkennung Methoden der Sprachtechnologie und der *Computer Vision* voraus; die automatische Gebärdensprachsynthese bedient sich Verfahren aus der Sprachtechnologie und der Computergrafik, und bei multimodalen Dialogsystemen kommen potenziell alle drei Disziplinen zusammen.

Eine Einschränkung sei in diesem Zusammenhang angebracht: Sprachtechnologie ist nie perfekt. Sie ist es einerseits aus sprachinhärenten Gründen nicht – menschliche Sprache ist in hohem Masse produktiv, und Ambiguitätspotenzial besteht auf verschiedenen linguistischen Ebenen –, und andererseits wurden bis anhin für einzelne Sprachtechnologien die technologischen Möglichkeiten noch nicht voll ausgeschöpft.

Mit dem Aufkommen neuer maschineller Lernverfahren – im Speziellen neuronaler Netze – löst sich die modulare Architektur von ICT-Anwendungen etwa im Bereich Gebärdensprache zusehends auf: In Zukunft werden Systeme, die Kommunikation zwischen gehörlosen gebärdensprachbenutzenden Menschen und hörenden Menschen ermöglichen, die Erkennung, Übersetzung und Synthese von Gebärdensprache nicht mehr nacheinander ausführen (wie in Abb. 2, Abb. 3 und Abb. 4 dargestellt). Vielmehr wird es möglich sein, einzelne Schritte zusammenzuführen und so direkt

- vom Gebärdensprach-Input zum Lautsprach-Output (unter Zusammenführung der Schritte Erkennung und Übersetzung),
- vom Lautsprach-Input zum Gebärdensprach-Output (unter Zusammenführung der Schritte Übersetzung und Synthese) oder
- vom Input in einer Gebärdensprache in den Output in einer anderen Gebärdensprache (unter Zusammenführung der Schritte Erkennung, Übersetzung und Synthese) zu gelangen (Ebling, Camgöz & Bowden, im Erscheinen).

Ausblickend ist ferner festzuhalten, dass gerade die Verbindung von Sprachtechnologie und Sonderpädagogik noch grosses Forschungs- und Entwicklungspotenzial birgt. Dabei gilt es stets abzuwägen, wo der Einsatz von Technologie sinnvoll ist, denn Technologie darf nicht zum Selbstzweck verkommen. Ein grosser Bedarf besteht darin, die Zielgruppen in den Forschungs- und Entwicklungsprozess einzubeziehen. Selbst für die Evaluation von Sprachtechnologiesystemen wurden bis anhin nur in seltenen Fällen Zielpersonen hinzugezogen, etwa bei der automatischen Textvereinfachung (Drndarevic & Saggion, 2012; Evans, Orassan & Dornescu, 2014; Feng, 2010; Yaneva, Temnikova & Mitkov, 2015) oder im Kontext von Legasthenie (Rello et al., 2013). Für den Einbezug dieser Personengruppen in den gesamten For-

schungs- und Entwicklungsprozess gibt es entsprechend noch weniger Beispiele. Hier herrscht grosser Nachholbedarf.

Literatur

- Beals, K., Dahl, D., Fink, R. & Linebarger, M. (2015). *Speech and Language Technology for Language Disorders*. Boston: De Gruyter.
- Bosseler, A. & Massaro, D. (2003). Development and evaluation of a computer-animated tutor for vocabulary and language learning in children with autism. *Journal of Autism and Developmental Disorders*, 33 (6), 653–672.
- Boyes Braem, P. (2001). A multimedia bilingual database for the lexicon of Swiss German Sign Language. *Sign Language & Linguistics*, 4 (1/2), 133–143.
- Coster, W. & Kauchak, D. (2011). *Simple English Wikipedia: A new text simplification task*. www.aclweb.org/anthology/P11-2117 [Zugriff am 09.10.2018].
- Drndarevic, B. & Saggion, H. (2012). *Towards Automatic Lexical Simplification in Spanish: An Empirical Study*. www.aclweb.org/anthology/W12-2202 [Zugriff am 09.10.2018].
- Ebling, S. (2013). *Evaluating a Swiss German Sign Language Avatar among the Deaf Community*. Chicago. www.zora.uzh.ch/id/eprint/85717/1/CAMERA_READY_sltat2013_submission_14.pdf [Zugriff am 09.10.2018].
- Ebling, S. (2016). Hier liegt maschinelle Übersetzung auf der Hand. MDÜ: *Fachzeitschrift für Dolmetscher und Übersetzer*, 1, 32–35.
- Ebling, S., Camgöz, N. & Bowden, R. (im Erscheinen). Use of new technologies in L2 sign language assessment. In T. Haug, W. Mann & U. Knoch (Eds.), *Handbook on Language Assessment across Modalities*. Oxford: OUP.
- Evans, R., Orassan, C. & Dornescu, I. (2014). *An evaluation of syntactic simplification rules for people with autism*. www.aclweb.org/anthology/W14-1215 [Zugriff am 09.10.2018].
- Feng, L. (2010). *Automatic Readability Assessment*. Unveröffentlichte Dissertation, City University of New York.
- Frieg, H., Muehlhaus, J., Ritterfeld, U. & Bilda, K. (2017). Assistive Technologien in der Dysarthrietherapie. Entwicklung des Trainingssystems ISi-Speech als Anwendungsbeispiel. *Forum Logopädie*, 31 (3), 10–15.
- Ganzeboom, M., Yılmaz, E., Cucchiarini, C. & Strik, H. (2016). An ASR-Based Interactive Game for Speech Therapy. In *Proceedings of the 7th Workshop on Speech and Language Processing for Assistive Technologies (SLPAT 2016)* (pp. 63–68). San Francisco: ACL.

- Hawley, M., Cunningham, S., Green, P., Enderby, P., Palmer, R., Sehgal, S. & O'Neill, P. (2013). A voice-input voice-output communication aid for people with severe speech impairment. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 21 (1), 23–31.
- Heimann Mühlenbock, K. (2013). *I see what you mean: Assessing readability for specific target groups*. Unveröffentlichte Dissertation, University of Gothenburg.
- Kipp, M., Heloir, A. & Nguyen, Q. (2011). Sign Language Avatars: Animation and Comprehensibility. In H. Vilhjálmsson, S. Kopp, S. Marsella, & K. Thorisson (Eds.), *Proceedings of the 11th International Conference on Intelligent Virtual Agents (IVA)* (pp. 113–126). Reykjavík: Springer.
- Kipp, M., Nguyen, Q., Heloir, A. & Matthes, S. (2011). *Assessing the Deaf User Perspective on Sign Language Avatars*. www.michaelkipp.de/publication/Kippetal11b.pdf [Zugriff am 09.10.2018].
- Klaper, D., Ebling, S. & Volk, M. (2013). *Building a German/Simple German Parallel Corpus for Automatic Text Simplification*. www.zora.uzh.ch/id/eprint/78610/ [Zugriff am 09.10.2018].
- Krause, M., Smeddinck, J. & Meyer, R. (2013). A Digital Game to Support Voice Treatment for Parkinson's Disease. In W. Mackay, S. Brewster & S. Bødker (Eds.), *CHI '13 Extended Abstracts on Human Factors in Computing Systems* (pp. 445–450). New York: ACM.
- Lehr, M., Prud'hommeaux, E., Shafran, I. & Roark, B. (2012). *Fully Automated Neuropsychological Assessment for Detecting Mild Cognitive Impairment*. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.386.4718&rep=rep1&type=pdf> [Zugriff am 09.10.2018].
- Liesen, C. & Rummler, K. (2016). Digitale Medien und Sonderpädagogik. Eine Auslegeordnung für die interdisziplinäre Verbindung von Medienpädagogik und Sonderpädagogik. *Schweizerische Zeitschrift für Heilpädagogik*, 4, 6–12.
- Lopez-Cozar, R., Callejas, Z., Espejo, G. & Griol, D. (2011). Enhancement of *Conversational Agents by Means of Multimodal Interaction*. In D. Perez-Marin & I. Pascual-Nieto (Eds.), *Conversational Agents and Natural Language Interaction: Techniques and Effective Practices* (pp. 223–252). Hershey: IGI.
- Martens, H., Nuffelen, G. V., Dekens, T., Huici, M. H.-D., Hernández-Díaz, H. A. K., Letter, M. D. & Bodt, M. D. (2015). The effect of intensive speech rate and intonation therapy on intelligibility in Parkinson's disease. *Journal of Communication Disorders*, 58, 91–105.
- Milne, M., Luerssen, M., Lewis, T., Leibbrandt, R. & Powers, D. (2011). Designing and Evaluating Interactive Agents as Social Skills Tutors for Children with Autism Spectrum Disorder. In D. Perez-Marin & I. Pascual-Nieto (Eds.),

- Conversational Agents and Natural Language Interaction: Techniques and Effective Practices* (pp. 23–48). Hershey: IGI.
- Newman, J. (2014). *To Siri, With Love*. www.nytimes.com/2014/10/19/fashion/how-apples-siri-became-one-autistic-boys-bff.html [Zugriff am 12.10.2018].
- Pompili, A., Amorim, C., Abad, A. & Trancoso, I. (2015). *Speech and language technologies for the automatic monitoring and training of cognitive functions*. www.inesc-id.pt/publications/11553/pdf [Zugriff am 09.10.2018].
- Prillwitz, S., Leven, R., Zienert, H., Hanke, T. & Henning, J. (1989). *HamNoSys: Version 2.0: Hamburger Notationssystem für Gebärdensprachen: Eine Einführung*. Hamburg: Signum.
- Rello, L., Bayarri, C., Gòrriz, A., Baeza-Yates, R., Gupta, S., Kanvinde, G., Saggi-
on, H., Bott, S., Carlini, R. & Topac, V. (2013). DysWebxia 2.0! More Acces-
sible Text for People with Dyslexia. [www.ims.uni-stuttgart.de/institut/mit-
arbeiter/bottsn/Publicat/w4a2013-Challenge-Dyswebxia.pdf](http://www.ims.uni-stuttgart.de/institut/mit-arbeiter/bottsn/Publicat/w4a2013-Challenge-Dyswebxia.pdf) [Zugriff am
09.10.2018].
- Romero-Fresco, P. (2011). *Subtitling through speech recognition: Respeaking*.
Manchester: St. Jerome.
- Rudzicz, F., Namasivayam, A. & Wolff, T. (2012). The TORGO database of
acoustic and articulatory speech from speakers with dysarthria. *Language
Resources & Evaluation*, 46 (4), 523–541.
- Rudzicz, F. (2016). *Clear Speech: Technologies that Enable the Expression and
Reception of Language*. San Rafael: Morgan & Claypool.
- Sagae, K., Lavie, A. & MacWhinney, B. (2005). *Automatic Measurement of
Syntactic Development in Child Language*. [www.cs.cmu.edu/~alavie/pa-
pers/sagae-aclo5-final.pdf](http://www.cs.cmu.edu/~alavie/papers/sagae-aclo5-final.pdf) [Zugriff am 09.10.2018].
- Smith, M., Ginger, E., Wright, K., Wright, M., Taylor, J., Boteler, H., Olsen, D.,
Bell, M. & Fleming, M. (2014). Virtual reality job interview training in
adults with autism spectrum disorder. *Journal of Autism and Developmental
Disorders*, 44, 2450–2463.
- Solorio, T. (2013). Survey on Emerging Research on the Use of Natural Lan-
guage Processing in Clinical Language Assessment of Children. *Language
and Linguistics Compass*, 7 (12), 633–646.
- Stein, D., Schmidt, C. & Ney, H. (2012). Analysis, preparation, and optimiza-
tion of statistical sign language machine translation. *Machine Translati-
on*, 26 (4), 325–357.
- Stoll, S., Camgöz, N., Hadfield, S. & Bowden, R. (2018). *Sign Language Pro-
duction using Neural Machine Translation and Generative Adversarial Net-
works*. <http://bmvc2018.org/contents/papers/0906.pdf> [Zugriff am
09.10.2018].

- Suendermann-Oeft, D. (2014). Modern Conversational Agents. In J. Jähnert & C. Förster (Hrsg.), *Technologien für digitale Innovationen*. Wiesbaden: Springer Fachmedien.
- Tartaro, A. & Cassell, J. (2008). *Playing with virtual peers: Bootstrapping contingent discourse in children with autism*. www.justinecassell.com/publications/TartaroCassell_ICLS_2008.pdf [Zugriff am 09.10.2018].
- Teodoro, G., Martin, N., Keshner, E., Shi, J. & Rudnick, A. (2013). Virtual clinicians for the treatment of aphasia and speech disorders. In *International Conference on Virtual Rehabilitation (ICVR)* (pp. 158–159). Philadelphia: ISVR.
- Vacher, M., Lecouteux, B., Istrate, D., Joubert, T., Portet, F., Sehili, M. & Chahua, P. (2013). *Experimental Evaluation of Speech Recognition Technologies for Voice-based Home Automation Control in a Smart Home*. www.slp-at.org/slp-at2013/W13/W13-3916.pdf [Zugriff am 09.10.2018].
- Veletsianos, G. & Russell, G. (2013). Pedagogical Agents. In J. Spector, M. Merrill, J. Elen & M. Bishop (Eds.), *Handbook of Research on Educational Communications and Technology* (pp. 759–771). New York: Springer.
- Yaneva, V., Temnikova, I. & Mitkov, R. (2015). *Accessible Texts for Autism: An Eye-Tracking Study*. www.qcri.org/app/media/4921 [Zugriff am 09.10.2018].

Dr. phil. Sarah Ebling
Senior Researcher (Dozentin)
Interkantonale Hochschule für Heilpädagogik Zürich
Schaffhauserstrasse 239
Postfach 5850
8050 Zürich
sarah.ebling@hfh.ch